STATE OF MICHIGAN

FILED

BEFORE THE MICHIGAN PUBLIC SERVICE COMMISSION

NOV -2 1998

MICHIGAN PUBLIC SERVICE COMMISSION

Ameritech Michigan's submission on performance measurements, benchmarks, and reporting in compliance with the October 2, 1998 Order in MPSC Case No. U-11654	))))	Case No. U-11830

AFFIDAVIT OF DANIEL S. LEVY

STATE OF ILLINOIS)
)
COUNTY OF COOK)

I, **DANIEL S. LEVY**, being duly sworn, state:

1. I have personal knowledge of the facts set forth herein, and I am competent to testify thereto as a witness.

I. Qualifications

2. My name is Daniel S. Levy. I have a Ph.D. in Economics from The University of Chicago. I serve as the National Leader of Economic Matters for Arthur Andersen's Strategy, Finance, and Economics Group. A copy of my resume may be found in Appendix B.

II. Purpose and Organization of the Affidavit

3. The purpose of this affidavit is to respond to the Commission's request for a simple yet rigorous methodology for analysis of Ameritech Michigan's performance measurements. The discussion below describes a methodology that will allow Ameritech Michigan, its competitors, and the Commission to determine whether Ameritech Michigan is meeting its contractual obligations. In order to make such a determination, Ameritech Michigan's performance in support of competitive local exchange carrier (CLEC) operations and end-users is either compared to performance standards or to Ameritech Michigan's performance in service of its retail customers. The results of these comparisons are used to determine whether or not Ameritech Michigan discriminates against CLECs in providing services to either the CLEC itself or its end-users.

4. Section III below details the proposed methodology. This methodology is based on statistical techniques that are well-known and accepted by courts, telecommunications companies, and academic experts. Section IV discusses the value of statistical techniques in assessing the quality of service that telecommunications companies provide in support of various groups of end-users. Section V develops the statistical measures that are used in the proposed methodology. Section VI details the proposed methodology for identifying when remedies will be employed. Section VII discusses the support that many

telecommunication companies have expressed for the use of statistical techniques similar to those proposed here. Section VIII provides conclusions.

III. Proposed Testing Method

5. This affidavit discusses two well-known statistical tests that can be used to determine whether Ameritech Michigan is providing non-discriminatory service. The testing protocol discussed is based on the z-test, which has been endorsed by a range of telecommunication companies and has been selected because it can be implemented through standard commercially available software. The test proposed employs a 95 percent, one-tailed, confidence interval. It calls for quarterly testing, which provides larger sample sizes that will increase the chance of identifying true disparity.

6. The method discussed here employs components of Ameritech's previous parity testing proposal to the FCC. At the request of the Michigan Commission Staff, Ameritech Michigan has simplified its previous proposal for parity testing. The parity testing methods proposed by Ameritech to the FCC do provide additional benefits that the Commission may want to consider. However, in order to comply with the wishes of the Commission Staff, Ameritech has proposed a simplified methodology.

7. An additional test, Fisher's exact test, is proposed for situations where the z-test is generally considered inappropriate due to specific characteristics of the measure being tested and comparatively small sample sizes. Again, the proposed test employs a 95 percent, one-tailed, confidence interval. Like the z-test, this test is well-known and generally accepted.

8. The proposed tests are based on quarterly data, which will increase sample sizes, and thus increase the ability to detect disparity of performance during the test period. In addition, as mentioned above, larger sample sizes yielded by quarterly testing will reduce the impact of random fluctuations in performance that are likely to result from random chance.

IV. Benefit of Statistical Methodology

9. Obviously, the level of performance experienced by Ameritech Michigan's own end-users will vary from quarter to quarter, month to month, and even from day to day. For each performance measure, a given result in a quarter contains a random component.[1] The observed performance of Ameritech Michigan on any given performance measure will change from one period to the next even if the underlying performance of Ameritech Michigan is consistent over time.

[1] See AT&T *ex parte* communication to the FCC dated February 3, 1998

Therefore, the observed performance of Ameritech Michigan in a given quarter is viewed as a "sample", in statistical terms, of the underlying level of performance provided to end-users by Ameritech Michigan.

10. Similarly, even though Ameritech Michigan may be providing equal levels of service to both its own and CLEC end-users, random variation and chance will result in differences in measured performance for CLEC and Ameritech Michigan transactions during any given measurement period. The statistical methods discussed here can be used to distinguish between differentials in performance generated by random chance and those attributable to Ameritech Michigan.

11. Because of the complexity of factors that affect Ameritech Michigan's performance, it is likely that on occasion these standard tests will indicate discrimination when in fact there is no discrimination. It is possible that more detailed analysis of the source of disparity may demonstrate that the appearance of disparity is erroneous. This additional analysis may require further levels of disaggregation or alternative statistical methods. In some cases, the apparent disparity will not reflect true disparate service, but rather will be attributable to some acceptable market factor that was not reflected in the first-stage analysis.

12. For example, consider the situation of a CLEC, which submitted, in March 1998, a disproportionate number of its 911 customer record updates on March 25th.

Because the computer systems were malfunctioning on that date, the customer record update processes were delayed for both CLEC and Ameritech Michigan retail customers. The data indicate that this CLEC received disparate treatment in March. However, the apparent disparity was not due to discriminatory service. If the CLEC had not submitted a disproportionate number of files on March 25th, its performance data would indicate parity for that month.

13. The statistical analyses and testing protocols that are outlined in this affidavit are based on the assumption that if parity is not observed, the first course of action should be to investigate whether there is an explanation for the apparent disparity.

14. The statistical methods outlined have the following goals.

- Provide a high likelihood of correctly assessing remedies for disparity when disparity exists.
- Provide a low likelihood of incorrectly assessing remedies when parity exists.
- Provide a comparison of performance that reduces the impact of random variation.
- Provide a testing protocol that is easy to implement and verify.

15. Statistical tests provide the ability to achieve these goals. In addition, statistical tests such as these have been recognized by regulators, courts, and the scientific

community. For these reasons among others, many telecommunications companies have agreed that statistical tests should be employed for performance testing.[2] The tests have been chosen because of their ease of implementation, which the Staff has requested. In some situations, it may be appropriate to employ alternative tests. However, the z-test and Fisher's exact test are a pair of well-known and accepted tests that can be employed in a broad range of relevant settings.

A. Why a Statistical Methodology is Necessary

16. Statistical tests are designed to measure whether observed differences in performance are unlikely to result from anything other than the typical random variation that would be expected in this type of data. Consider the situation in which Ameritech Michigan provides exactly identical repair service to both CLEC and retail customers. In any quarter, the observed service to CLEC and retail customers will be slightly different due to random variations in the types of problems that occur. To the extent that these differences are small, they may not reflect a meaningful difference between Ameritech Michigan's measured performance for CLEC and retail end-users.

[2] See Section VI below.

17. For example, in March 1998, the average time required to provide a repair — Mean-Time-to-Repair — for Ameritech Michigan retail customers was 39.32 hours. Even if Ameritech Michigan were providing non-discriminatory service, it is unlikely that the Mean-Time-to-Repair for a CLEC's resale customers would be exactly 39.32 hours. Instead, because of random variation in performance for all customers, service to CLEC customers will be worse than service to retail customers about half of the time. Of course, the other half of the time performance to retail customers will be worse than that provided to CLEC customers. It will almost never be the case that performance to any two groups of end-users will be exactly the same.

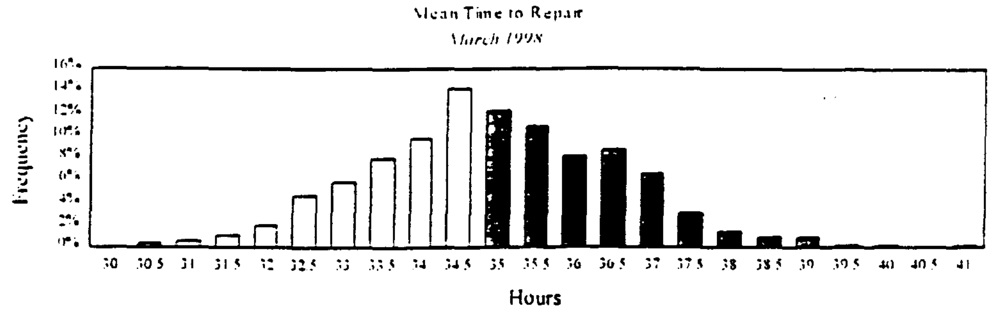
18. The effect of this type of random variation on observed performance is not unique to the telecommunications industry. It also affects many aspects of our everyday lives, from the complex to the mundane. Consider, for example, a perfectly fair coin that is tossed 500 times. One expects to see the coin come up heads 250 times and tails 250 times. But in fact this does not always happen. There is more than a 16-percent chance that the 500 tosses will result in more than 261 heads. There is an equal probability that the 500 tosses will result in less than 239 heads. These results follow from the laws of probability. If one concludes that the coin is biased based on a criterion of observing at least 261 heads, a fair coin would be erroneously judged as being biased 16 percent of the time.

19. To make the coin toss example more directly comparable to Ameritech Michigan's situation, consider the following scenario. Two fair coins are tossed 500 times each. We know that both coins will tend to produce 250 heads. A non-statistical methodology might conclude that there was disparity whenever the "CLEC" coin produced fewer heads than the "Ameritech Michigan retail" coin. However, even though both coins are fair, due to random variation there is a low probability that both coins will produce the exact same number of heads in a set of 500 tosses. A non-statistical methodology that judged the "parity of performance" of the two coins based simply on whether one coin produced more heads than the other would indicate an apparent disparity nearly 50 percent of the time, even though the two coins were both perfectly fair. Clearly, two fair coins are in exact parity all of the time. It is simply random variation that leads to apparent disparity half of the time.

20. This same type of random variation affects Ameritech Michigan's observed performance in its service to any two randomly selected groups of end-users in a given month or quarter. Consider a specific example drawn entirely from Ameritech's own retail customers' experience with Mean-Time-to-Repair. To demonstrate the effects of random chance on measured performance, I randomly selected 1,000 groups, each containing 1,000 end-users, from Ameritech's retail customer base. Because these selected end-users are actual Ameritech customers, they are receiving an underlying level of service that is, by definition,

nondiscriminatory compared to

Figure 1
Apparent Discrimination Caused by Random Chance



the broader pool of Ameritech customers, which experienced a Mean-Time-To-Repair of 34.7 hours.[3] The graph above illustrates the variation in Mean-Time-to-Repair for these groups.

21. As Figure 1 shows, about 48 percent of the 1,000 groups sampled had a Mean-Time-to-Repair above the average of 34.7 hours. If one were to judge performance and award remedies on a non-statistical basis, Ameritech Michigan would make remedy payments to approximately 483 out of 1,000 groups *of its own retail customers.*

22. Statistical methodologies recognize the inherent variability in the type of performance data at issue here. The steps needed to conduct these statistical tests

[3] Note that this calculation includes trouble reports that came clear, and trouble reports with no trouble found. Also, the calculation reflects the average across all of Ameritech's retail customers, not just Michigan.

are described in greater detail below.

V. Basic Statistical Concepts and Terms

A. Binary Data versus Continuous Data

23. There are two broad categories of data reflected in Ameritech Michigan's performance measurements: binary data and continuous data. Binary measures have only two possible outcomes for a given event. For instance, the Trouble-Report-Rate is a binary measure, since there are only two possibilities for a given phone line: either it had a trouble, or it didn't. Similarly, Confirmed-Due-Dates-Not-Met is a binary measure since for any particular due date there are only two options: either the due date was met, or it was not met.

24. In contrast, continuous data can take on any value along a continuum. For instance, Mean-Time-To-Repair is a continuous measure because the amount of time could be one minute, two days, or any other amount of time. Similarly, the Average-Installation-Interval is a continuous measure, exhibiting a wide range of possible values for the measured amount of time to install.

B. The Mean

25. A primary question in performance measurement is whether Ameritech Michigan's "typical" service for CLEC end-users is different from the "typical" performance provided to Ameritech Michigan's own retail customers. The "mean," or "average," is a widely used measure of typical performance, representing the "center" of a group of values. The mean can be interpreted as the "expected value" of the data. Clearly, some customers will experience a longer repair time than the mean, and some will experience a shorter repair time, but the mean repair time provides an "expected" length of time that a customer will tend to wait for a repair.

26. The mean of a group of values, or the sample mean, is simply the sum of all of the observed data divided by the number of observations. It is important that each observed value be included in the calculation. For instance, if a certain CLEC had 10 trouble reports in a given quarter, and 7 took 1 hour to repair, and 3 took 3 hours to repair, the Mean-Time-to-Repair would be 1.6 hours. The mean reflects both the values of the data *and* the frequency with which those values are observed.

$$\text{Mean} = \bar{x} = \frac{\Sigma x_i}{n}, \text{ where}$$

x_i indicates the observed values of the data, where x_1 is the first observed value, x_2 is the second value, etc.
Σ indicates the summation. In this case, all of the observed values are summed.
n is the total number of observed values.

27. The same calculation can be used for both continuous and binary measures. The only distinction that arises with binary measures is that each outcome is assigned a value of either zero or one. For instance, with Confirmed-Due-Dates-Not-Met, each due date is assigned the value zero if the due date was met, or assigned the value one if the due date was not met. If there were 100 installations in a given quarter, and 95 of the confirmed due dates were met and 5 were missed, then the mean of this measure would be equal to 0.05, or 5 percent, which is equal to the sum of the values of the due date data (5) divided by the total number of observations (100).

28. Ameritech Michigan's mean observed performance supplied to any group of end-users can be calculated and compared to the observed performance supplied to any other group. As discussed above, due to random variations, it will be rare that Ameritech Michigan's observed performance will be identical for CLEC and Ameritech Michigan end-users. However, with parity of service, it will also be rare to observe large differences in performance between CLEC and Ameritech Michigan end-users.

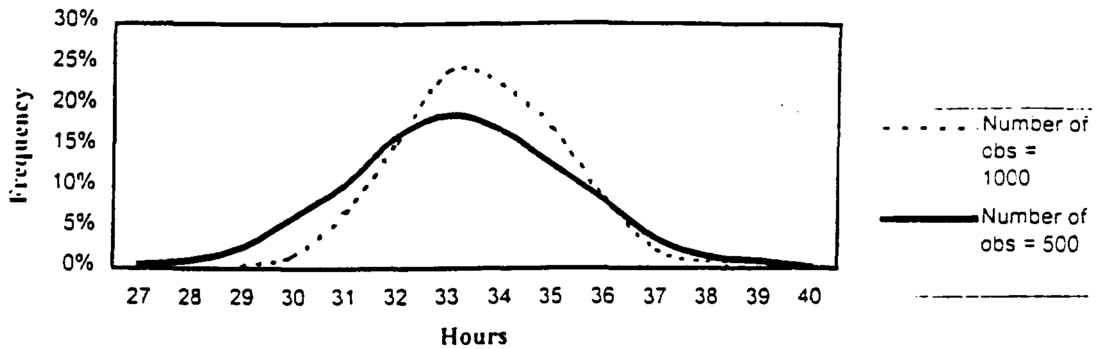
C. Variance and Standard Deviation

29. The larger the sample size used to determine the mean, the less variation there will be in the observed mean. When a mean is based on very few observations, there is a risk that a single extreme value will have a large impact on the estimated mean. A mean calculated with many observations will be less susceptible to the influence of a single or small number of extreme values.

30. The effect of sample size on the variability of the means can be seen in Figure 2. The dotted line reflects the distribution of the means from groups with 1,000 end-users, while the solid line reflects the distribution of the mean when the sample includes 500 end-users. Again, the information for this graph has been drawn from the actual experience of Ameritech's retail customers. The dotted line reflects the greater precision of the estimate of the mean that is achieved with larger sample sizes. Notice that proportionally fewer of the groups of 1,000 observations are at the more extreme values, above approximately 36 and below 32. More of the groups of 1,000 observations are found close to 34.7, the mean for Ameritech's retail customers for March 1998. With more observations, we are more likely to obtain a group that has a mean closer to the true mean of the underlying population.

Figure 2: The Effects of Number of Observations on the Distribution of the Monthly Mean

Mean-Time-to-Repair - *March 1998*



31. Since, as Figure 2 demonstrates, the estimate of the mean is more precise when the group is larger, it is possible to identify smaller differences in means between two groups, when the groups are larger. The statistical tests described below specifically account for the fact that larger groups provide a more precise estimate, increasing the probability of detecting disparity when it exists.

32. Of course, the precision of the estimated mean is not only influenced by the sample size, it is also affected by the amount of variation in the underlying measure. If there is little or no variation in a measure, its mean may be estimated very precisely with relatively little data. Measures that exhibit large variation in the performance experienced by individual end-users will require larger sample

sizes to achieve the same level of precision.

33. In order to determine whether the observed differences in performance are likely to result merely from random chance, we need a measure of the variability of the performance measure across CLEC and Ameritech Michigan end-users that reflects two things: a) the variation in the underlying performance data, and b) the number of observations. The variance of the mean and the standard deviation of the mean are common statistical measures of variability in data. With these measures of variability, it is possible to determine whether observed differences in mean performance levels between groups are likely to result solely from random chance.

34. The variance of the data is calculated as follows. The first step is to subtract the mean of the data set, $\bar{x}$, from each observation in the data set, x_i. These individual differences from the mean are squared and summed. This total squared difference is divided by the number of observations minus one to create a measure of the dispersion in the data.[4]

$$\text{Variance of the sample} = \frac{\Sigma[x_i - \bar{x}]^2}{(n-1)} = \sigma^2 \text{, where } \sigma \text{ is pronounced "sigma"}$$

[4] A variance based on data samples is a sample or estimated variance and is more correctly referred to as s^2. The population variance is referred to as σ^2. However, most documentation of parity testing in the telecommunications context has referred to sample variances as σ^2, and we will use this convention as well unless otherwise noted. The notational difference is not important, but the conceptual difference is.

Standard deviation of the sample = $\sqrt{\sigma^2}$ = square root of the variance = σ

35. These fundamental statistical concepts are the basis of a well-known statistical test known as the z-test. In the next section, I describe the use of the z-test in implementing the proposed parity test.

V. Proposed Statistical Tests

A. The z-test

36. Ameritech has proposed the use of the z-test to determine whether there is a statistically significant difference between the mean level of performance provided to two groups. As discussed above, the goal of statistical testing is to achieve a high probability of awarding remedies when there is true disparity, while reducing the probability that remedies will be awarded when performance is not in parity. Achieving these two outcomes will depend on the dispersion of the underlying data for the measure in question for both the CLEC and Ameritech Michigan, as well as the number of observations in the quarter.

37. The z-test is based on an index for comparing measurement results from different sources of data. The index is based on the difference between two means. In this

case, the index is based on the difference between the mean performance for CLEC customers and the mean performance for Ameritech Michigan retail customers. The difference between the two measures is simple to compute.

$$\text{DIFF} = \bar{x}_{AIT} - \bar{x}_{CLEC}$$

38. The z-test index adjusts the difference between the two means based on the standard deviation of that difference. As discussed above, the standard deviation measures the dispersion of the data and provides a threshold for the typical variation in the data. The standard deviation of the difference between the means depends on the variance of the performance for CLEC customers, the variance of the performance for retail customers, and the number of observations of CLEC and retail performance data.[5]

$$\text{Variance}_{DIFF} = \sigma^2_{DIFF} = \frac{\sigma^2_{CLEC}}{n_{CLEC}} + \frac{\sigma^2_{AIT}}{n_{AIT}} = \text{variance of difference between the means}$$

$$\text{Standard deviation}_{DIFF} = \text{sqrt}(\sigma^2_{DIFF}) = \sigma_{DIFF} = \text{standard deviation of difference between the means}$$

39. The z-test can be one-tailed or two-tailed. The one-tailed version of the z-test identifies cases of disparity in one direction, in this case when the service to the

[5] See Appendix A to this affidavit for a more detailed discussion.

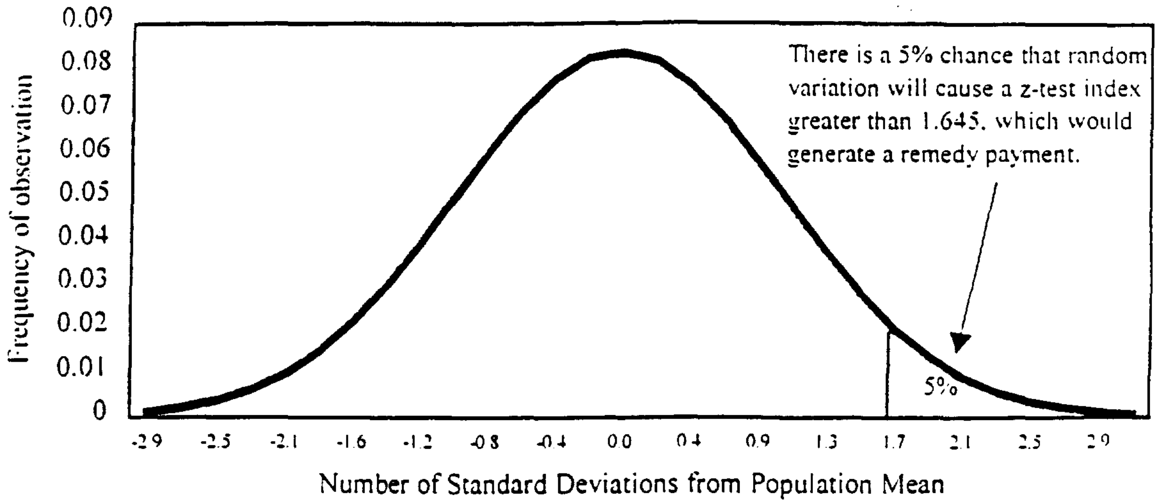
CLEC end-users is worse than that provided to Ameritech end-users. The two-tailed version identifies disparity in either direction. Since the object of this statistical analysis is to test whether Ameritech Michigan's service provision in its resale market is worse than it is for its retail customers, the one-tailed z-test is more appropriate than the two-tailed version.

40. The z-test that Ameritech Michigan has proposed will tend to produce a finding of disparity 5 percent of the time even when parity exists. This means that in situations where Ameritech Michigan is in parity, it will tend to pay a penalty 5 percent of the time, even though it provides the same level of service to both CLEC and Ameritech Michigan end-users. This finding of disparity will occur simply due to the random variations in the data. This is depicted in Figure 3, which shows the probability of observing sample means that are at increasing distances from the population mean. The z-test is designed so that there is only a 5 percent chance that the sample mean will be more than 1.645 standard deviations above the population mean.

41. At the same time, the z-tests proposed here will detect significant levels of disparity when they exist. For example, based on data from March 1998, a difference of as little as 2.4 hours in Average-Installation-Interval between Ameritech Michigan and AT&T end-users would be detected by the proposed test 98 percent of the time and would be defined as disparity. This means that the z-

test we have proposed has a high likelihood of detecting differences in the level of performance when they exist.

Figure 3
Distribution of Sample Means from Population Mean



42. As is typical with statistical tests, the ability to detect differences in performance will depend on how many observations are available for the test. Therefore, for measures that do not occur as frequently as installations, larger differences would have to occur before they would become statistically significant. Similarly, differences in performance between CLECs and Ameritech Michigan will become more likely to be detected as the size of CLECs increases. This means that if CLECs increase in size over time, the statistical tests proposed here will become increasingly strict, requiring more similar levels of performance between the CLECs and Ameritech Michigan before service would be considered in parity.

43. Typically, the conditions for employing a z-test for the comparisons of the mean of two continuous measures requires at least 30 observations.[6] Therefore, Ameritech Michigan has proposed that tests of parity be employed only in those situations where both CLEC and Ameritech Michigan end-users have more than 30 observations for a given performance measure being tested.

44. Ameritech has proposed that whenever the z-test index is less than 1.645 performance would be considered "in parity." For parity tests where the z-test index is greater than 1.645, the measure would be considered out of parity, unless Ameritech could demonstrate that more appropriate disaggregation levels or alternative statistical tests were more appropriate for the specific circumstances of that measure in the given quarter. As stated above, Ameritech has proposed quarterly testing, which will increase sample sizes, facilitating the identification of true disparity.

45. Although the z-test is a valid and acceptable test for parity for measures that are calculated as proportions, the underlying assumptions of the z-test are not valid when data sets are small and the proportions tend to be extreme (close to one or zero). The required minimum sample size for the z-test depends on the observed

[6] See Appendix A for additional discussion on required sample sizes.

proportion in the data (see Table 1 below).[7] Because some of the measures of interest in this setting (for instance, Confirmed-Due-Dates-Not-Met, or Trouble-Report-Rate) exhibit small proportions (sometimes less than one percent) and some of the carriers have small sample sizes, in some cases it may be more appropriate to use the Fisher's exact test. In addition to its merits when data sets are small and observed probabilities are low, the Fisher's exact test is also valid when the sample sizes are larger and the observed probability is closer to 50 percent. However, we are recommending the z-test, when it is appropriate, because of its additional power. The calculations for Fisher's exact test are described in detail in the Appendix A. They can be implemented on standard, commercially available computer software.

Table 1
Implementation of the z-test
Required Minimum Sample Sizes at
Different Levels of Observed Probabilities[8]

Sample Proportion	Sample Size
0.5	≥ 30
0.4 or 0.6	≥ 50
0.3 or 0.7	≥ 80
0.2 or 0.8	≥ 200
0.1 or 0.9	≥ 600
0.05 or 0.95	≥ 1400

[7] See further discussion in the Appendix A.
[8] Zar (1984), pp. 385–386.

VII. CLECs have endorsed the use of a statistical methodology

46. There is wide recognition in the telecommunications industry that statistical methods are essential for measuring parity of service. The use of statistical methodology has been endorsed by many of Ameritech Michigan's competitors in the local exchange markets.

A. Endorsements of the use of a statistical methodology

- The following statement has been excerpted from a document created by the Local Competition Users Group (LCUG), a cooperative effort of AT&T, MCI, Sprint, LCI and WorldCom, dated February 6, 1998 (version 1.0), page 4.

  When making the comparison of ILEC results to CLEC results, it is necessary to employ comparative procedures that are based upon generally accepted statistical procedures. It is important to use statistical procedures because all of the ILEC-CLEC processes that will be measured are processes that contain some degree of randomness. Statistical procedures recognize that there is measurement variability, and assist in translating results data into useful decision-making information.

- "Allegiance [Allegiance Telecom, Inc.] agrees that statistical analysis is an essential tool in determining whether or not an ILEC is meeting its obligation to provide competing carriers with nondiscriminatory interconnection and access to OSS, operator services and directory assistance." Source: Comments of Allegiance Telecom, Inc., dated June 1, 1998. CC Docket No. 98-56, RM-9101, p. 7.

- "A statistically valid method to evaluate parity is critical to the overall performance requirement process. Parity cannot be fairly determined without an appropriate statistical methodology." Source: Comments of MCI Telecommunications Corporation, dated June 1, 1998. CC Docket No. 98-56, RM-9101, p. iii.

- "Sprint agrees with the Commission that reporting averages of performance measurements alone may not suffice in uncovering underlying differences in performance. Thus, Sprint supports the use of statistical techniques for determining whether there are statistically significant differences between the ILEC's performance when provisioning service to its own retail customers and its performance toward competing carriers." Source: Comments of Sprint Corporation, dated June 1, 1998. CC Docket No. 98-56, RM-9101, p. 6.

B. Endorsements of the z-test

47. The z-test methodology is consistent with standard methods of statistical testing and has been endorsed by many telecommunications companies.[9] The following statements indicate their support for this methodology:

- "MCI and the Local Users Group have recommended a statistical methodology called 'the z test'. After examining various statistical tests, LCUG members determined that the 'z test' methodology best adjusts for the probability of errors (1) pointing to parity violations where none exists and (2) missing parity violations where they do exist." MCI, Pennsylvania CLEC/ILEC facilitation, p. 10, III.
- "With respect to the statistical test, the PUCT [Public Utilities Commission of Texas] has approved the Z-test to determine the parity of a performance measurement in SWBT's interconnection agreements with AT&T and MCI." Public Utility Commission of Texas, NPRM Comments, p. 8.

VII. Conclusions

50. Ameritech has recommended a statistical methodology as a test for parity of service. The statistical methodology provides Ameritech Michigan with strong

incentives to maintain nondiscriminatory performance to its CLEC customers by prescribing a substantial level of remedies wherever there is apparent discrimination. Statistical methods adjust for the day-to-day random variation in the data, distinguishing between performance differences that may be evidence of discrimination and performance differences that could arise due to random chance. They provide a standard, generally accepted methodology for identifying apparent discrimination.

51. While the statistical methods described above can be used to compare performance between Ameritech Michigan's resale and retail markets, they do not test for discriminatory behavior. Rather, they indicate how likely it is that differences in the service provided to each market are or are not due to random chance. Findings of apparent disparity would not necessarily indicate discriminatory intent or behavior on the part of Ameritech Michigan. Further analysis based on statistical tests that are more appropriate for a specific situation may reject the existence of disparity. Moreover, additional investigation of particular circumstances for a particular measure for a given quarter may demonstrate the disparity was not the result of Ameritech Michigan's actions, but rather was caused by conditions beyond Ameritech Michigan's control.

" See, for example, the LCUG response to the NPRM for the 271 rulemaking.

52. It should also be noted that even instances of correctly identified disparity may not justify a legal conclusion of discrimination. either because the cause of the disparity is beyond the control of Ameritech Michigan or because the magnitude of the disparity does not warrant such a conclusion. As noted by the FCC, "...even if statistically significant differences appear between results for the incumbent LEC and the competing carrier, these differences may be too small to have any practical competitive consequence and may not justify a legal conclusion that the incumbent LEC has discriminated against the competing carrier."[10]

[10] Notice of Proposed Rulemaking; Appendix B — Statistical Methods. p. B4.

This concludes my affidavit.

Further affiant saith not.

Daniel S. Levy

Subscribed and sworn

before me this 31 day of

October, 1998.

COMM. EXPIRES 01/11/2001

APPENDIX A

I. Appropriate Statistical Methods for Testing Parity: Form of Data and Distributional Assumptions

The statistical analysis proposed by Ameritech Michigan compares Ameritech Michigan's performance in providing service to its own retail customers with its performance in providing service to customers of its competing carriers. Findings of inferior service provision to its competing carrier customers would indicate "disparity" between Ameritech's retail and resale markets. Otherwise, Ameritech Michigan's performance in both markets would be considered "in parity." The statistical methods used to test for parity often depend on the form of the data — binary or continuous — describing each of the performance measures being examined.

Binary data are classified into two discrete categories. For example, whether a line was or was not installed in time is a binary outcome. For binary data, we compare the frequency of such occurrences for retail customers versus competing carrier customers. For example, if the proportion of lines needing repairs is 2 percent of all retail lines but 1 percent of all a competing carrier's lines, a statistical test could determine how likely it is that this difference is due to random chance.

Continuous data measure a quantity or a length (e.g., *how long* it took to repair a line, rather than whether a line did or did not need a repair). A comparison of means is often appropriate for performance measures based on continuous data. For example, if the average time needed to repair a line is two days for the retail market but three days for the competing carrier or resale market, a statistical test could determine how likely it is that this hypothetical difference is due to random chance.

Calculation of the probability that an observed difference is due solely to random chance depends on the assumptions made regarding the distribution of the data. Choice of the appropriate test, therefore, depends on making the appropriate distributional assumptions given the data to be analyzed. Whether the data are binary or continuous often is important, if not determinative, in making these assumptions. I discuss these issues below as they apply to parity testing.

I.A. Statistical Tests for Performance Measures Based on Continuous Data

i) Pooled vs. Separate Variance Tests

When comparing means of continuous performance measures, the z-test is employed.[1] Two different versions of the test may be used depending on the assumptions made about the variance, or spread, of the populations from which the means were sampled. If the variance of the populations from which both the retail and the resale data were sampled are assumed to be equal, a z-test using a combined or pooled variance estimate may be used.[2] Otherwise, separate variance estimates from the retail and resale sample data are used.[3] I currently use the separate variance version of the z-test in the comparisons of means since this test makes the fewest assumptions about the underlying populations. Statistical comparisons of variances should not be considered part of a z-test but rather, require a different statistical test altogether.[4] Should such tests reveal systematic equality in the variances of the retail and resale markets, the pooled version of the z-test can be used if and where appropriate.

ii) One-tailed v. Two-tailed Tests

The z-test can be one-tailed or two-tailed. The one-tailed version of the test only identifies cases of disparity in one direction — either resale performance being worse than retail, or vice versa. The two-tailed version identifies disparity in either direction. Since the object of this statistical analysis is to test whether Ameritech's service provision in its resale market is worse than in its retail market, the one-tailed z-test is more appropriate than the two-tailed version.

iii) Sample Size

To obtain accurate results when comparing means using a z-test, the means must be distributed according to a normal distribution. According to the central limit theorem, as sample size increases, the distribution of sample means becomes increasingly normal. This result holds for virtually all distributions of data. As one increases sample size, the speed with which the distribution of sample means approaches normality depends on how closely the underlying population from which the data were sampled follows a normal distribution. Sample sizes of 30 observations are commonly viewed as a minimum threshold for the distribution of sample means to approach normality.[5] Currently, a sample of 30 observations is used as a minimum sample size in the parity tests proposed by Ameritech Michigan, for both samples of CLEC and Ameritech Michigan end-users, when comparing means using a z-test. This threshold can be increased if warranted by indications of significant departures from normality in the data.

I.B. Statistical Tests for Performance Measures Based on Binary Data

i) Limitations of the z-test comparing proportions

Binary data follow a binomial distribution.[6] For large sample sizes with sample proportions close to 0.5, the binomial distribution converges to a normal distribution. Under such circumstances, a z-test can be used to compare differences in proportions.[7] However, the smaller the sample size or the more the sample proportions deviate from 0.5, the less appropriate the assumption of normality. Many of the performance measures based on binary data in the present case involve comparisons of proportions of less than 5 percent, and sometimes less than 1 percent. In addition, the sample sizes are often well under recommended levels for using the normal approximation.[8] Due to these data limitations, there may be situations where Fisher's exact test is more appropriate.

ii) Fisher's exact test

Fisher's exact test is a widely understood and generally accepted statistical test for comparing proportions that can be used when sample sizes are small or sample proportions are close to zero or one. This test does not require a minimum sample size or restrict its application to a limited range of sample proportions because it is an exact statistical test. It does not rely on an approximated distribution, but rather calculates the exact probability of obtaining specific frequencies of observations. The simple example below is illustrative.

For the performance measure comparing the percentage of lines needing repairs, the observed frequencies of retail and resale lines can be arranged in a 2x2 ("two-by-two") table (see Table 1). The two columns identify retail or resale observations, and the two rows identify lines needing or not needing repairs. When testing for parity, the null hypothesis is that the percentage of resale lines needing repairs is equal to or less than the percentage of retail lines needing repairs. In this example, if Fisher's exact test indicates a statistically significant difference between 25.0 percent (the resale repair rate) and 23.1 percent (the retail repair rate), we would reject the null hypothesis in favor of the alternative hypothesis — that the percentage of resale lines needing repairs is greater than the percentage of retail lines needing repairs.

First, the probability of obtaining the observed frequencies is calculated.[9] Then, for fixed row and column totals, the corresponding probabilities of every other 2x2 table which is "more extreme" than the observed 2x2 table are calculated and summed. For the two-tailed version of the test, more extreme tables are those which are less likely than the observed table. For the one-tailed version of the test, which is more appropriate for this analysis, more extreme tables are those indicating worse resale performance than the observed table (Tables 2-4 below).[10]

This sum of probabilities is added to the probability of obtaining the observed table, yielding a p-value, which is the result of the test. A large p-value (close to one) would indicate a high probability of obtaining the observed difference under the null hypothesis, and a low p-value (close to zero) would indicate a low probability of obtaining the observed difference under the null hypothesis. Comparing the p-value to a pre-determined level of statistical significance, typically set at α=0.05, determines whether or not the p-value is small enough to indicate disparity.

Table 1 – Observed Frequencies
Probability = 0.431

	Resale	Retail	Total
No Repair	3	60	63
Repair	1	18	19
Total	4	78	82
% Repair	25.0%	23.1%	

Table 2 – More Extreme Table
Probability = 0.191

	Resale	Retail	Total
No Repair	2	61	63
Repair	2	17	19
Total	4	78	82
% Repair	50.0%	21.8%	

Table 3 – More Extreme Table
Probability = 0.035

	Resale	Retail	Total
No Repair	1	62	63
Repair	3	16	19
Total	4	78	82
% Repair	75.0%	20.5%	

Table 4 – More Extreme Table
Probability = 0.002

	Resale	Retail	Total
No Repair	0	63	63
Repair	4	15	19
Total	4	78	82
% Repair	100.0%	19.2%	

The one-tailed Fisher's exact test above yields the p-value:
p = 0.431 + 0.191 + 0.035 + 0.002 = 0.659. If the pre-determined level of statistical significance were set at α=0.05, we would clearly fail to reject the null hypothesis that the proportion of resale lines needing repairs was the same as the proportion of retail lines needing repairs because the p-value is much greater than α. In other words, this comparison of proportions indicates parity between the retail and resale markets for this performance measure.[11]

The example above, together with Table 5 below, demonstrate how random chance and sample size affect determinations of parity. A more superficial treatment of the data which, for example, compared absolute percentage differences to test for parity might make a determination of disparity in the case above. However, given the relatively small sample size, Fisher's exact test indicates that it is highly likely that the observed difference is just due to chance. If the same proportions were observed but with much larger sample sizes, the likelihood that a resale repair rate 1.9 percentage points greater than the retail repair rate was due solely to chance would be greatly reduced. We observe this situation in table 5 below.

Table 5 – Alternate Observed Frequencies
Probability = 0.000093

	Resale	Retail	Total
No Repair	3,750	461,400	465,150
Repair	1,250	138,600	139,850
Total	5,000	600,000	605,000
% Repair	25.0%	23.1%	

The p-value for a Fisher's exact test on the alternate frequencies observed in Table 5 is p=0.00087, indicating that with a larger sample size, it is very unlikely that a difference as large as 1.9 percentage points would be observed if the null hypothesis of equal population proportions was true. Since this p-value is less than the pre-determined level of statistical significance of α=0.05, we would reject the null hypothesis of parity.

These examples indicate the necessity of using a statistical approach in parity analyses. The choice of the appropriate test and recognition of the influence of sample size are required to correctly account for randomness in data. Use of non-statistical approaches when comparing means and proportions would often lead to errors of interpretation due to the statistical uncertainty in sampled data.

II. Other Statistical Tests

The statistical methods described above are appropriate to apply when testing for parity given: a) the data and performance measures being examined to date. and b) the Staff's request for a simple method of implementing a test for parity. If additional performance measures require examination. additional data become available, or further analysis reveals the need to reexamine the methods that have been applied to date. the appropriate application of other statistical methods may prove useful. Some of these methods include Bayesian tests. which allow for the incorporation of prior beliefs about the data. Others include nonparametric tests. bootstrapping. and permutation tests. which are subject to limited. if any. constraints regarding distributional assumptions.

References

Agresti, Alan. 1990. *Categorical Data Analysis*. New York: John Wiley & Sons.

Cochran, William G. 1977. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons. Cited in Jerrold H. Zar, *Biostatistical Analysis*, 2nd ed. (Englewood Cliffs, NJ: Prentice Hall, 1984), pp. 385–386.

Evans, Merran, Nichols Hastings, and Brian Peacock. 1993. *Statistical Distributions*. 2nd ed. New York: John Wiley & Sons.

Fisher, R.A. 1935. "The Logic of Inductive Inference." *Journal of the Royal Statistical Society*, Ser. A. 98: 39–54.

Fisher, R.A. 1934. *Statistical Methods for Research Workers*. 5th ed. Edinburgh, Scotland: Oliver and Boyd.

Kmenta, Jan. 1986. *Elements of Econometrics*. 2nd ed. New York: Macmillan Publishing Company.

Larsen, Richard J., and Morris L. Marx. 1986. *An Introduction to Mathematical Statistics and Its Applications*. 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Lancaster, H.O. 1961. "Significance Tests in Discrete Distributions." *J. American Statistical Association* 45: 223–234. Cited in Alan Agresti, *Categorical Data Analysis* (New York: John Wiley & Sons, 1990), p. 66.

Little, Roderick J. A. 1989. "Testing the Equality of Two Independent Binomial Proportions." *The American Statistician* 43: 283–288.

Martín Andrés, A., and I. Herranz Tejedor. 1995. "Is Fisher's Exact Test Very Conservative?" *Computational Statistics & Data Analysis* 19: 579–591.

Martín Andrés, A., I. Herranz Tejedor, and A. Silva Mato. 1995. "The Wilcoxon, Spearman, Fisher, X^2, Student and Pearson Tests and 2 x 2 Tables." *The Statistician* 44: 441–450.

Matlack, William F. 1980. *Statistics for Public Policy and Management*. North Scituate, MA: Duxbury Press.

Notice of Proposed Rulemaking, Federal Communications Commission, CC Docket No. 98-56, RM-9101, FCC 98-72.

Storer, Barry E., and Choongrak Kim. 1990. "Exact Properties of Some Exact Test Statistics for Comparing Two Binomial Proportions." *Journal of the American Statistical Association* 85, no. 409: 146–155.

Wonnacott, Thomas H., and Ronald J. Wonnacott. 1984. *Introductory Statistics for Business and Economics*. 3rd ed. New York: John Wiley & Sons.

Yates, F. 1984. "Tests of Significance for 2 X 2 Contingency Tables." *J. Royal Statist. Soc.*, Ser. A, 147, Part 3: 426–463.

Zar, Jerrold H. 1996. *Biostatistical Analysis*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Zar, Jerrold H. 1984. *Biostatistical Analysis*. 2nd ed. Englewood Cliffs, NJ: Prentice Hall

[1] The score of a z-test is the difference in the sample means relative to the standard deviation of this difference. The standard error is a measure of the spread of the data that provides an estimate of the typical deviation of a difference in sample means from zero. The null hypothesis for these parity tests — that the population means are equal — is expressed mathematically by assuming the difference in population means is zero. The larger the observed difference in means for a fixed standard error, the larger the score of the z-test and the more likely it is that the sample means are indeed "different," or obtained from two different populations rather than from the same underlying population.

In this analysis our z-test calculations use estimated variances from the sample data. When estimated variances are used, the results of the test follow a student's t distribution (Wonnacott and Wonnacott, 1984, p. 264). If sample sizes are large, however, and the estimated variances can be assumed to be the population variances, the student's t distribution will approximate the normal distribution (Wonnacott and Wonnacott, 1984, p. 264). In this analysis, we base statistical inferences made from the scores of the z-test on the normal distribution. From Ameritech's perspective, this is a conservative approach to testing for parity, because the kurtosis of the normal distribution is smaller than that of the student's t distribution (Zar, 1996, p. 95). This makes findings of disparity more likely than if we relied on the student's t distribution. The probability density functions of both the normal and student's t distributions are listed below (Larsen and Marx, 1986):

$$if\ X \sim N(\mu,\sigma),\ then\ f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} \times e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{(p. 210)}$$

where μ = population mean
σ = population standard deviation
π = mathematically defined constant ≈ 3.14159
e = mathematically defined constant ≈ 2.71828

$$if\ X \sim t(n),\ then\ f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \times \Gamma\left(\frac{n}{2}\right) \times \left(1+\frac{x^2}{n}\right)^{\left(\frac{n+1}{2}\right)}} \qquad \text{(p. 341)}$$

where n = degrees of freedom

Γ = gamma function where $\Gamma(r) = \int_0^\infty x^{(r-1)} e^{-x} dx$ (p. 227).

[2] The formula for the pooled-variance z-test is (Zar, 1996, p. 125):

$$Z = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

where $\overline{X_1}$ = incumbent LEC sample mean
$\overline{X_2}$ = CLEC sample mean

μ_1 = incumbent LEC population mean

μ_2 = CLEC population mean

$(\mu_1 - \mu_2) = 0$ under null hypothesis of parity, or equality of means

$$s_p^2 = \text{pooled sample variance} = \frac{\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1\right)^2 + \sum_{i=1}^{n_2}\left(X_{2i} - \overline{X}_2\right)^2}{(n_1 - 1) + (n_2 - 1)},$$

(Wonnacott and Wonnacott, 1984, p. 232)

where X_{1i} = each incumbent LEC observation $(i = 1,2,3...)$ and X_{2i} = each CLEC observation $(i = 1,2,3...)$

n_1 = incumbent LEC sample size

n_2 = CLEC sample size

If statistical inferences based on the result of the test assume a student's t distribution rather than the normal distribution (a t-test instead of a z-test), the degrees of freedom for this test is:
$df = (n_1 - 1) + (n_2 - 1)$ (Wonnacott and Wonnacott, 1984, p. 232).

[3] The formula for the z-test which does not assume equal variances is (Kmenta, 1986, p. 137 and p. 145):

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $s_1^2 = \frac{\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1\right)^2}{n_1 - 1}$ = incumbent LEC sample variance (Matlack, 1980, p.47)

$s_2^2 = \frac{\sum_{i=1}^{n_2}\left(X_{2i} - \overline{X}_2\right)^2}{n_2 - 1}$ = CLEC sample variance (Matlack, 1980, p.47).

This version of the z-test is an approximate solution. This problem, know as the Behrens-Fisher problem, has remained unsolved for over 50 years (Larson and Marx, p.362).
If statistical inferences based on the result of the test assume a student's t distribution rather than the normal distribution (a t-test instead of a z-test), the approximation for the degrees of freedom for this test is (Zar, 1996, p. 129):

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2 - 1)}}.$$

[4] One test of equality of variances of two assumed normal populations is given by the formula below (Kmenta, 1986, pp. 147-148):

$$\frac{s_1^2}{s_2^2} \sim F_{n_1-1,n_2-1}$$

where F_{n_1-1,n_2-1} is the F distribution with $n_1 - 1, n_2 - 1$ degrees of freedom. This tests the null hypothesis that $\sigma_1^2 \leq \sigma_2^2$ against the alternative hypothesis that $\sigma_1^2 > \sigma_2^2$ where σ_1^2 and σ_2^2 are the two population variances.

[5] If the populations from which the two means are sampled are normal *and* the variances of these populations are identical, the z-test is an exact test. Consequently, it is not subject to the sample size constraints imposed by reliance on approximations to the normal distribution based on the central limit theorem.

[6] The formula for the probability function of the binomial distribution is (Zar, 1996, p. 515):

$$if\ W \sim B(n, p), then\ p_W(x) = \frac{n!}{x!(n-x)!} p^x q^{(n-x)}$$

where n = sample size
p = probability of event occurring
k = number of events occurring.

[7] The formula for the test statistic comparing differences in proportions from large sample sizes is (Martín Andrés, Herranz Tejedor, and Silva Mato, 1995, p.444):

$$Z = \frac{p_1 - p_2 - (\pi_1 - \pi_2)}{\sqrt{p_{both}(1 - p_{both})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where p_1 = sample incumbent LEC proportion = $\frac{k_1}{n_1}$ = events in sample / sample size

p_2 = sample CLEC proportion = $\frac{k_2}{n_2}$ = events in sample / sample size

$$p_{both} = \frac{k_1 + k_2}{n_1 + n_2}$$

π_1 = population LEC probability

π_2 = population CLEC probability

$(\pi_1 - \pi_2) = 0$ under null hypothesis of parity, or equality of proportions

n_1 = incumbent LEC sample size

n_2 = CLEC sample size

Like the z-test on means, the z-test above compares the difference between two proportions relative to the standard error of the difference of these sample proportions. It, too, is based on the assumption of normally distributed means, because proportions are means for binary data. This assumption again allows us to determine the statistical confidence with which we can say the sample proportions are the "same," or drawn from the same underlying binary distribution.

[8] Zar (1984, pp. 385–386) provides the following table from Cochran (1977, p.58) with sample size recommendations for different magnitudes of sample proportions:

Sample proportion	Sample Size
0.5	≥ 30
0.4 or 0.6	≥ 50
0.3 or 0.7	≥ 80
0.2 or 0.8	≥ 200
0.1 or 0.9	≥ 600
0.05 or 0.95	≥ 1,400

[9] The formula for obtaining the probability of any specific 2x2 table is (Zar, 1996, p. 541):

$$P = \left(\frac{R_1!R_2!C_1!C_2!}{n!}\right)\left(\frac{1}{f_{11}!f_{12}!f_{21}!f_{22}!}\right)$$

where R_1 and R_2 = row 1 total and row 2 total, respectively

C_1 and C_2 = column 1 total and column 2 total, respectively

f_{11} = count in cell: row 1, column 1

f_{12} = count in cell: row 1, column 2

f_{21} = count in cell: row 2, column 1

f_{22} = count in cell: row 2, column 2

$n = f_{11} + f_{12} + f_{21} + f_{22}$ = total number of observations

Fisher's exact test is based on combinatorics. Since the row and column totals are fixed, as one of the four cell counts varies, the other three are adjusted accordingly and the probability of observing each resulting 2x2 table follows the hypergeometric distribution given by the formula (Evans, Hastings, and Peacock, 1993, p. 85):

$$\text{if } V \sim H(N, X, n), \text{ then } p_V(x) = \frac{\binom{X}{x}\binom{N-X}{n-x}}{\binom{N}{n}}$$

where N = total sample size (total number of lines in the example in the body of the Appendix A)
X = number of events in total sample (total number of lines needing repair)
n = sample size of comparison category (total number of retail lines)
x = number of events in comparison category (number of retail lines needing repair)
where the minimum value of x is $\max(0, n - N + X)$ and the maximum value of x is $\min(X, n)$.

[10] As with the z-test comparing means, since the object of this statistical analysis is to test whether Ameritech's service provision in its resale market is worse than it is for its retail customers, the one-tailed version of Fisher's exact test is more appropriate than the two-tailed version.

[11] An arguable limitation of Fisher's exact test is that it conditions on both the row and column marginal totals, meaning that both the row and column totals must be fixed during the calculation. Although fixing the row totals (repair / no repair) is an unnecessary restriction for testing the resale and retail proportions being compared, many statisticians have argued that this does not significantly detract from the accuracy of the test, and they do not hesitate to advocate its use.

"Fisher's exact test is the most widely known and accepted method for analyzing a 2x2 table ..." (Andrés Martín and Herranz Tejedor, 1995, p.590).

"It is the probabilities of occurrence in the relevant subset that provide the correct basis for tests of significance. In other words, we must condition on the margins, whatever the origin of the table. Whether no, one, or two margins are 'fixed' in advance is irrelevant." (Yates, 1984, p. 433)

Fisher (1935) and later Yates (1984, together with many other discussants (including Barnard and Cox), argue that knowledge of the joint distribution of the row totals provides little inference on the magnitude of association evident in a 2x2 table. Little (1989, p.286) describes the row marginal sums as "approximately ancillary" because little information is lost by conditioning on both marginals.

"The Fisher exact test is applicable to contingency tables where both the row totals and column totals are set in advance of data collection (an uncommon situation). Fortunately, the testing procedure appears to work with other contingency tables as well" (Zar, 1984, p. 392).

One should note, especially when working with small samples, that Fisher's exact test is conservative. This decreases the power of the test and increases the likelihood of making Type II errors — accepting the null hypothesis of parity when parity does not exist. However, it also reduces the probability of making Type I errors — rejecting the null hypothesis of parity when parity does, in fact, exist. Statisticians have argued that the benefits of Fisher's exact test outweigh any reduction in power resulting from its use over other methods "... the loss of power produced by using Fisher's test is very slight in the majority of situations, and this is acceptable in return for the greater ease of computation and a more generic validity (for all types of sample)." (Andrés Martín and Herranz Tejedor, 1995, p.579).

Some statisticians have proposed adjustments to Fisher's exact test to compensate for its conservatism. Agresti (1990) cites Lancaster's (1961) and Plackett's (discussion of Yates, 1984) advocacy of the mid-P value method — using half the P value of the observed table plus the probability of the more extreme tables — "as a good compromise between having a conservative test and using randomization on the boundary to eliminate problems from discreteness" (Agresti, 1990, p.66).

Appendix B

DANIEL S. LEVY ECONOMIST

EDUCATION

Ph.D., Economics, University of Chicago
A.B., Economics, University of Chicago (With Special Honors in Economics)

Daniel S. Levy specializes in applications of economics and statistics in the study of corporate structures related to industrial organization/antitrust and transfer-pricing issues. His work includes detailed analyses and valuations of corporate functions, risks, and assets for international corporations in a wide range of industries. Dr. Levy's work also includes the study of environmental issues, including comment before the EPA on contingent valuations of power plant emissions damages. In the area of labor economics, he has studied the effects of variations in employment incentives on the productivity and retention of employees, and has investigated the social and economic determinants of investments in human capital. He is expert in numerous statistical and modeling applications, and has modeled complex economic and social factors affecting demographic and market behavior.

Prior to joining Arthur Andersen, Dr. Levy held research and consulting positions at Charles River Associates, The RAND Corporation, Needham-Harper Worldwide Advertising, SPSS Inc. and the University of Chicago Computation Center.

EXPERT TESTIMONY/AFFIDAVITS

- Before the FCC, 1998, Expert Affidavit, *Statistical Analysis.*
- Graber, A. et al. v. Giuliani, United States District Court Southern District of New York, 1998, Expert Affidavit and Deposition, *Statistical Sampling and Survey Research.*
- Marisol, A. et al. v. Giuliani, United States District Court Southern District of New York, 1998, Expert Affidavit and Deposition, *Statistical Sampling and Survey Research.*
- DFW v. Continental Air Lines, Texas, 1998, Expert Deposition and Testimony.
- Randall's Food Markets, Inc., v. Fleming Companies, Inc., The American Arbitration Association Dallas, Texas, June, 1998, Expert Affidavit, *Statistical Sampling.*
- Randall's Food Markets, Inc., v. Fleming Companies, Inc., The American Arbitration Association Dallas, Texas, February 1998, Expert Report, *Statistical Sampling.*

- Donald E. Haney v. Timesavers Inc., et al. United States District Court, District of Oregon, January 1998, Expert Testimony, *Patent Infringement.*

- Merck-Medco Managed Care Inc. v. Rite Aid Corporation et al. Northern District of Maryland, May 1997, Expert Deposition, *Antitrust.*

- Donald E. Haney v. Timesavers Inc., et al. United States District Court, District of Oregon, July 1997, Expert Report, *Patent Infringement.*

- Kenneth Heubert Williams v. Honri Vashon Hunt et al., State of Michigan in the Circuit Court for the County of Oakland, May 1997, Expert Deposition, *Value of Life.*

- Merck-Medco Managed Care Inc. v. Rite Aid Corporation et al. Northern District of Maryland, April 1997, Expert Report, *Antitrust.*

- Robinson Rubber et al. V. Hennepin County, Minnesota, United States District Court, District of Minnesota, Fourth Division, April 1997, Expert Deposition, *Antitrust.*

- Robinson Rubber et al v. Hennepin County, Minnesota, United States District Court, District of Minnesota, Fourth Division, April 1997, Expert Report, *Antitrust.*

- Massachusetts Wholesalers of Malt Beverages, Inc., v. Commonwealth of Massachusetts et al, Suffolk Superior Court, 1996, Expert Testimony, *Financial Damages.*

- Luke Brothers v. S. P. Krusell, US District Court, District of Massachusetts, July 1996, Expert Affidavit, *Antitrust.*

- Luke Brothers v. S. P. Krusell, US District Court, District of Massachusetts, August 1996, Expert Affidavit, *Antitrust.*

- Daras v. Texaco Inc, 1993, Affidavit.

- Environmental Protection Agency: Navajo Generating Station, 1991, Public Comment, *Valuation of Environmental Damages.*

PROFESSIONAL EXPERIENCE

1996 – Present	Director of Economics, Arthur Andersen L.L.P.:CRCO
1995 - 1996	Economist, Arthur Andersen L.L.P.
1991 - 1995	Senior Associate, Charles River Associates

1988 - 1991	Associate Economist. The RAND Corporation
1985 - 1988	Computer Advisor. University of Chicago Computation Center
1982 - 1985	Research and Teaching Consultant, SPSS Inc.
1981 - 1982	Research Consultant. Needham, Harper Worldwide Advertising

PROFESSIONAL HONORS AND ACTIVITIES

- Earhart Fellowship for graduate research in economics. 1981 - 1982
- Hewlett Grant for research in developing countries, 1985 - 1986; renewed, 1986 - 1987
- CBS Bicentennial Scholarship for research on events leading to the American Revolution, 1986 - 1987
- Homer and Alice Jones Fellowship, University of Chicago, 1987 - 1988
- American Economics Association, 1988- Present
- Population Association of America, 1988-1991

PAPERS, PRESENTATIONS, AND PUBLICATIONS

Daniel S. Levy. "New Econometric Techniques for Transfer Pricing." Presented at the American Bar Association Annual Meetings, August, 1997.

Daniel S. Levy et al. "Economics and the New Transfer Pricing Regulations: Achieving Arm's Length Through the Invisible Hand." Special Report to *Transfer Pricing Reporter*, Vol. 4, No. 2, May 24, 1995.

Daniel S. Levy and Deloris R. Wright. "In the OECD and the United States, It's the Arm's-Length Principle that Matters: Comparison of New Transfer Pricing Regulations." *International Transfer Pricing Journal* 1, No. 2, January 1995.

Robert Fagan, Manjusha Gokhale, Daniel S. Levy, Peter Spinney, and G.C. Watkins. "Estimating DSM Program Impacts for Large Commercial and Industrial Electricity Users." Presented at 1995 International Energy Program Evaluation Conference, Chicago, IL, August 1995.

Talk on the EPA's decision to require the Navajo Generating Station to reduce emissions to protect visibility in the Grand Canyon. Panel on "Valuation of Environmental Resource Damages," CRA conference on *Economists' Perspectives on Legal Issues Today: Estimating Damages*, Boston, MA, April 23, 1992.

Daniel S. Levy et al. "Conceptual and Statistical Issues in Contingent Valuation: Estimating the Value of Altered Visibility in the Grand Canyon." (MR-344-RC). Santa Monica, CA: RAND Corporation, 1995. Draft submitted to the Environmental Protection Agency, March 1991.

Daniel S. Levy and D. Friedman. "The Revenge of the Redwoods?: Reconsidering Property Rights and Economic Allocation." *The University of Chicago Law Review* (April 1, 1994). Reprinted in *Land Use and Environment Law Review* 26 (September 1995).

Lois Davis, Susan Hosek, Daniel S. Levy and Janet Hanley, "Health Benefits for Military Personnel: An Overview of Their Value and Comparability to Civilian Benefits" (WD-5875-FMP). Santa Monica, CA: RAND Corporation, February 1992.

D. Buddin, J. Hanley, Daniel S. Levy, and D. Waldman. *Promotion Tempo and Enlisted Retention* (R-4135-FMP). Santa Monica, CA: RAND Corporation, August 1991.

Daniel S. Levy et al. "Comments On Contingent Valuation of Altered Visibility in the Grand Canyon Due to Emissions from the Navajo Generating Station." Presented to the Environmental Protection Agency, April 18, 1991.

Daniel S. Levy. "The Economic Demography of the Colonial South." Ph.D. Thesis, Department of Economics, University of Chicago, 1991.

J. DaVanzo and Daniel S. Levy. "Influences on Breastfeeding Decisions in Peninsular Malaysia." Presented at *The Yale Conference on the Family, Gender Differences, and Development*, September 1989.

Daniel S. Levy. "Long-Run Geographic and Temporal Changes in Mortality in the Colonial South." Presented at the annual meeting of the Population Association of America, Baltimore, 1989. Submitted 1995 to *Social Science History*.

Daniel S. Levy. "The Economic Determinants of Family Sizes in Colonial Maryland: Evidence from Colonial Legislators of Maryland." Presented at the Social Science History Association, Chicago, 1989.

Daniel S. Levy. "The Epidemiological Causes of Changing Political Life Expectancies." Manuscript, 1989.

Daniel S. Levy. "The Life Expectancies of Colonial Maryland Legislators." *Historical Methods* 20, No. 1 (Winter 1987): 17–27.

David W. Galenson and Daniel S. Levy. "A Note on Biases in the Measurement of Geographic Persistence Rates." *Historical Methods* 19, No. 4 (Fall 1986): 171–179.